Appendix

Appendix A

Data Cleaning Procedures

The data cleaning process involved several steps to ensure the datasets were ready for analysis:

1. **Handling Missing Values:**

o **ISTAT Data:** Missing values in healthcare expenditures were imputed using the median value of the available data.

o **Ministry of Health Data:** Missing prevalence rates of Long COVID were interpolated based on nearby time points.

o **OECD Data:** Missing economic indicators were filled using forward-fill and backward-fill methods.

2. **Normalization:**

o All monetary values were adjusted to 2024 euros using inflation rates provided by ISTAT.

o Labor force participation rates were normalized to a scale of 0 to 1 for consistent analysis.

3. **Data Merging:**

o The datasets from ISTAT, Ministry of Health, OECD, and European Commission were merged based on common time indices (years).

o Care was taken to align the fiscal years of different data sources to avoid temporal mismatches.

4. Outlier Detection and Handling:

o Outliers were detected using Z-scores. Data points with Z-scores greater than 3 were considered outliers and were handled by capping at the 99th percentile.

Assumptions

1. **Pessimistic Scenario:** Assumes a 5% increase in Long COVID cases and a 2% increase in healthcare spending, indicating a severe scenario with minimal intervention.

2. **Neutral Scenario:** Assumes a 3% increase in Long COVID cases and a 1% increase in healthcare spending, reflecting a balanced approach.

3. **Optimistic Scenario:** Assumes a 1% increase in Long COVID cases and a 0.5% increase in healthcare spending, representing an effective intervention scenario.